Effort Scores for Family Planning Programs: An Alternative Approach

John A. Ross and Katharine Cooper-Arnold

July 2000

MEASURE Evaluation

Carolina Population Center
University of North Carolina
at Chapel Hill
123 W. Franklin Street
Suite 304
Chapel Hill, NC 27516
Phone: 919-966-7482
Fax: 919-966-2391
measure@unc.edu
www.cpc.unc.edu/measure

Collaborating Partners:

Macro International Inc.
11785 Beltsville Drive
Suite 300
Calverton, MD 20705-3119
Phone: 301-572-0200
Fax: 301-572-0999
measure@macroint.com

John Snow Research and Training Institute
1616 N. Ft. Myer Drive
11th Floor
Arlington, VA 22209
Phone: 703-528-7474
Fax: 703-528-7480
measure_project@jsi.com

Tulane University
1440 Canal Street
Suite 2200
New Orleans, LA 70112
Phone: 504-584-3655
Fax: 504-584-3653
measure2@tulane.edu

Funding Agency:

Center for Population, Health
and Nutrition
U.S. Agency for
International Development
Washington, DC 20523-3600
Phone: 202-712-4959

WP-01-28

The research upon which this paper is based was sponsored by the MEASURE *Evaluation* Project with support from the United States Agency for International Development (USAID) under Contract No. HRN-A-00-97-00018-00.

The working paper series is made possible by support from USAID under the terms of Cooperative Agreement HRN-A-00-97-00018-00. The opinions expressed are those of the authors, and do not necessarily reflect the views of USAID.

The working papers in this series are produced by the MEASURE *Evaluation* Project in order to speed the dissemination of information from research studies. Most working papers currently are under review or are awaiting journal publication at a later date. Reprints of published papers are substituted for preliminary versions as they become available. The working papers are distributed as received from the authors. Adjustments are made to a standard format with no further editing.

A listing and copies of working papers published to date may be obtained from the MEASURE *Evaluation* Project at the address listed on the back cover.

Other MEASURE *Evaluation* Working Papers

WP-00-27 Monitoring Quality of Care in Family Planning Programs: A Comparison of Observation and Client Exit Interviews (Ruth E. Bessinger and Jane T. Bertrand)

WP-00-26 Rating Maternal and Neonatal Health Programs in Developing Countries (Rodolfo A. Bulatao and John A. Ross)

WP-00-25 Abortion and Contraceptive Use in Turkey (Pinar Senlet, Jill Mathis, Siân L. Curtis, and Han Raggers)

WP-00-24 Contraceptive Dynamics among the Mayan Population of Guatemala: 1978-1998 (Jane T. Bertrand, Eric Seiber and Gabriela Escudero)

WP-00-23 Skewed Method Mix: a Measure of Quality in Family Planning Programs (Jane T. Bertrand, Janet Rice, Tara M. Sullivan & James Shelton)

WP-00-22 The Stymied Contraceptive Revolution in Guatemala (Roberto Santiso G. and Jane T. Bertrand)

WP-00-21 The Impact of Health Facilities on Child Health (Eric R. Jensen and John F. Stewart)

WP-00-20 Effort Indices for National Family Planning Programs, 1999 Cycle (John Ross and John Stover)

WP-00-19 Evaluating Malaria Interventions in Africa: A Review and Assessment of Recent Research (Thom Eisele, Kate Macintyre, Erin Eckert, John Beier, Gerard Killeen

WP-00-18: Monitoring the AIDS epidemic using HIV prevalence data among young women attending antenatal clinics: prospects and problems (Basia Zaba, Ties Boerma and Richard White)

WP-99-17: Framework for the Evaluation of National AIDS Programmes (Ties Boerma, Elizabeth Pisani, Bernhard Schwartländer, Thierry Mertens)

WP-99-16: National trends in AIDS knowledge and sexual behaviour in Zambia 1996-98 (Charles Banda, Shelah S. Bloom, Gloria Songolo, Samantha Mulendema, Amy E. Cunningham, J. Ties Boerma)

WP-99-15: The Determinants of Contraceptive Discontinuation in Northern India: A Multilevel Analysis of Calendar Data (Fengyu Zhang, Amy O. Tsui, C. M. Suchindran)

WP-99-14: Does Contraceptive Discontinuation Matter?: Quality of Care and Fertility Consequences (Ann Blanc, Siân Curtis, Trevor Croft)

WP-99-13: Socioeconomic Status and Class in Studies of Fertility and Health in Developing Countries (Kenneth A. Bollen, Jennifer L. Glanville, Guy Stecklov)

WP-99-12: Monitoring and Evaluation Indicators Reported by Cooperating Agencies in the Family Planning Services and Communication, Management and Training Divisions of the USAID Office of Population (Catherine Elkins)

WP-98-11: Household Health Expenditures in Morocco: Implications for Health Care Reform (David R. Hotchkiss, Zine Eddine el Idriss, Jilali Hazim, and Amparo Gordillo)

WP-98-10: Report of a Technical Meeting on the Use of Lot Quality Assurance Sampling (LQAS) in Polio Eradication Programs

WP-98-09: How Well Do Perceptions of Family Planning Service Quality Correspond to Objective Measures? Evidence from Tanzania (Ilene S. Speizer)

WP-98-08: Family Planning Program Effects on Contraceptive Use in Morocco, 1992-1995 (David R. Hotchkiss)

WP-98-07: Do Family Planning Service Providers in Tanzania Unnecessarily Restrict Access to Contraceptive Methods? (Ilene S. Speizer)

WP-98-06: Contraceptive Intentions and Subsequent Use: Family Planning Program Effects in Morocco (Robert J. Magnani)

WP-98-05: Estimating the Health Impact of Industry Infant Food Marketing Practices in the Philippines (John F. Stewart)

WP-98-03: Testing Indicators for Use in Monitoring Interventions to Improve Women's Nutritional Status (Linda Adair)

WP-98-02: Obstacles to Quality of Care in Family Planning and Reproductive Health Services in Tanzania (Lisa Richey)

WP-98-01: Family Planning, Maternal/Child Health, and Sexually-Transmitted Diseases in Tanzania: Multivariate Results using Data from the 1996 Demographic and Health Survey and Service Availability Survey (Jason Dietrich)

Effort Scores for Family Planning Programs: An Alternative Approach

by

John A. Ross and Katharine Cooper-Arnold

July 2000

John A. Ross is a Senior Fellow, and Katharine Cooper-Arnold is a Research Associate, both with The Futures Group International, 80 Glastonbury Blvd., Glastonbury, CT 06033.
Email J.Ross@TFGI.com

Effort Scores for Family Planning Programs: An Alternative Approach

Abstract:

In the 1999 cycle of the Family Planning Program Effort Scores the standard questionnaire was used as before to produce 30 indices of effort. However a shorter form was also included, which asked the respondent to rate each of the 30 indices directly on a 1-10 scale. Overall the total short score averaged about 8 percent below the total long score, consistently across 88 countries, geographic regions, and among score components. The different score levels appear to arise not from those sources but rather from differences in the two instruments. The long form produces the 30 indices from about 125 questions, using various questions and coding rules to create each score, whereas the short form described each index in brief phrases and used a simple scale for each one. Most differences were in fact relatively small, and especially so where the two forms were most similar. For a consistent time trend for each score in each country it will be useful to retain the long form, but for the total score and the component scores the short form could now be used at less cost and perhaps more frequently to gauge program effort.

Acknowledgements: This research was funded by The MEASURE Project, University of North Carolina. We thank John Stover for important suggestions for the analysis, and Kate Abel for assistance with data processing

Effort Scores for Family Planning Programs: An Alternative Approach

by John A. Ross and Katharine Cooper-Arnold

Nearly 30 years ago Lapham and Mauldin (1972) invented measures for the strength of large-scale family planning programs and applied them to twenty developing countries. Ten years later, in 1982, they expanded the list of measures and developed a detailed questionnaire, which they sent to a large number of observers of programs throughout the developing world. That study was replicated in 1989, 1994, and 1999, producing indices that are meant to capture program inputs, independent from such outputs as increased contraceptive use or fertility change (Mauldin and Ross, 1996; Ross and Mauldin, 1997; Ross and Stover, 2000). These have been widely used by researchers, as they constitute the only series that applies to most countries, at the same point in time, with results on 30 program features presented in simple scales. Many of the 30 items concern program features for which empirical data are generally unavailable, so judgmental questions are answered by respondents who are closely familiar with each country program. A test in two countries that gathered intensive empirical data corresponding to each judgmental question found close agreement between the two (Mauldin et al., 1995).

This paper reports on a comparison of two approaches to these family planning program effort indices (termed the "FPE" scores). One approach is the "long questionnaire," used in the 1982-1999 cycles, which contains about 125 items that are reduced by conversion rules to the final set of 30 indices. The other approach, used also in 1999, is a much shorter form that simply lists the 30 indices with a brief description of each, and asks the respondent to rate effort on each one from very weak to very strong on a 1-10 scale.

The use of both the long form and the short form in the 1999 round was meant to determine whether the two forms would produce similar results for the overall picture of program effort, and at what level of detail. Particular questions concerned how country rankings on the total and the component scores would differ

between the short and the long scores, how well the short scores would predict the long scores, and whether the questionnaire could be shortened without loss of information and without disturbing the long-term time trend.

The impetus to test a short version of the FPE questionnaire originated with the EVALUATION Project at the University of North Carolina. A simple version should reduce the cost and time needed to mount the study and might let rounds be done more frequently. The main FPE questionnaire places fairly heavy demands upon the respondents and has been used only about every five years. A more frequent series might strengthen the monitoring of family planning programs around the world. Another consideration was to test the construct reliability of the main version through a second approach.

However certain constraints apply to the promise of a short form of the questionnaire, since costs and time are saved only on parts of the survey operation. Some 750 respondents in about 100 countries must still be identified, and good addresses and contact numbers must be obtained. The questionnaires must be printed and mailed, and a follow-up system mounted. However, some savings should occur during the 3-6 months of follow-up since a short form is easier for respondents and they will probably fill it out quicker, and the response rate should be better. Further, data entry is simpler and faster, and the reconciliation of responses from multiple respondents in each country is far easier. All this would save personnel time and cost. (The reconciliation process would be quite mechanical since the short form uses a simple 10 point scale for each of the 30 scores and the respondent simply circles a number from one to 10, so respondents' replies would normally just be averaged. Reconciliation of replies from the long questionnaire usually follows the average also, but occasionally requires judgment.)

A preliminary test of the short form as originally designed was conducted by the EVALUATION Project at the University of North Carolina in Colombia, Paraguay, and Tanzania in 1998. The long form was implemented simultaneously and was scored by the same computerized conversion rules that were standard in the main international study. Respondents were randomly assigned either the short or long version to answer first. The results were promising: after converting both sets of scores to a common scale the long score ran no more than 10% above the short score (total FPE score) in each of the three countries.

That was the background to the decision to test the short form on a larger scale in the 1999 round of the international study, covering about 90 countries. Several options were considered. One was to drop the long form and use only the short form, since based upon the pretest the two seemed close enough that the results might turn out to be much the same as if the long form were used. The objections to this were that the pretest had involved only three countries, that the differences in results were not trivial, and that it was vital to protect the 22-year trend line in the overall FPE score for each country. Therefore it was decided to keep the long form and use it in all countries, as before.

Then the question was how to incorporate the short form. After considering various options, such as doing the comparison in half of the countries, or among half of the respondents in all countries, it was decided to give both the long and short forms to all respondents in all countries. A main object of the exercise was to learn whether the short form could be substituted for the long form in the future, and that required as robust a comparison as possible, by maximizing the number of comparisons within each country and involving all countries rather than half. The short form was placed at the end of the long form; in appearance the short form looks much different, and it was preceded by a brief introduction that asked for an additional picture of program effort.

However use of both forms lengthened the document and risked some damage to the response rate. As a partial counter-balance, staff were assigned to contact each respondent repeatedly in a systematic follow-up effort. This system used email for the first time (it was much less prevalent in 1994), as well as fax and telephone messages. However telephone calls overseas netted very little; email was by far the most useful. This follow-up effort had to be done separately in English, French, and Spanish and was executed with some interruptions due to staff turnover. The final response rate was not much different in 1999 than in previous rounds.

A further concern was that the quality of information would suffer due to respondent fatigue. Since a good time trend was paramount, the long form was placed first to protect it. Thus any respondent fatigue probably affected the short form more. In fact response was good to the short form. For 26 of the 30 scores,

92% -96% of respondents to the long form also filled out most or all of the short form. For the other four scores 86% - 90% of respondents filled out the short form.

Results

The analyses that follow assess the amount of agreement in the results from the two versions of the questionnaire. If agreement is close at all levels the briefer version might substitute for the longer version in future cycles of the study. In any case, the comparison is expected to clarify the nature of both instruments through an examination of the results for each of the 30 scores individually. The main results are presented below in order of detail: first the long-short overall difference, then differences by the four components, then differences by region, and then differences for each of the 30 items. The long form produced scores that range from zero to four, while the short form used a 1-10 scale (Appendix A). Both were converted to percent of maximum for comparability in some results below; in other cases the long form results were kept in the 0-4 scale and the short form results were converted to a 0-4 scale.

Overall: For the 88 countries in the analysis, the long total score averaged 53% of maximum possible, and the short total score averaged 47%. The six-point gap may be regarded as either small or large depending upon one's perspective; however it would have been important had the short form been used alone, since the time trend would have been flat in the five years since Cairo instead of rising by eight points in the long form results. The very consistent pattern of the higher long total score across the 88 countries appears in the Figure 1 scattergram. Most countries fall above the line of equality. However there is rather little scatter; a least squares line through the data points yields an R2 value of 0.76. Thus the long scores are higher, by fairly consistent amounts, from one country to another. That is the overall picture; the next question is how well it holds up at the more detailed levels.

By Four Components of the Total Score: One way of dividing the data is by the score components. The 30 scores are organized into four groups, related to policy, services, evaluation, and method availability. Again, a regular pattern emerges: the average long score is above each of the four short scores, rather evenly (Figure 2).

By Region: the long-short differences in the total score for all countries are confirmed within each region. That is, no region departs very far from the all-country pattern. Figure 3 shows the overall similarity; the differences appear about the same in all regions. Exact figures are in Table 1, which shows the overall short score at 89% of the overall long score, with little regional variation (next to last column).

By Both Regions and Components: Table 1 also shows the average scores for each of the four components. Two regions contain rather few countries (East Asia and the Central Asian Republics), which may help explain some of the differences.

By the 30 Indices: For each of the 30 indices the average long-short difference was obtained. Since each score could range from zero to four, the maximum difference was four; however the actual differences were close to l.0 or less; 20 of the 30 items are below 0.5 (Figure 4).

Figure 5 portrays this in a different way. Most differences are in favor of the long scores; only six items favor the short side and three of these are near zero. This says that the average figures found above, which favor the long scoring, are generally consistent across the 30 items and are not due to extreme values on a few of them. As a check, the countries were divided into two halves for the best and worst programs, those with the highest and lowest long total score, and the exercise was repeated. The patterns were essentially identical in both sets (not shown).

Distributions: The above comparisons use mean differences for all countries. The next step is to examine the distributions of the country differences through scattergrams. This is one approach to the question of how well the short form responses could predict the long form responses, and whether the short form might replace the long one. Those questions are examined now in their most demanding form, with reference to each item, for each country. An illustrative range of outcomes appears in Figures 6A-6C: Figure 6A for the Female Sterilization Score, where agreement was close; Figure 6B for Training, which heavily favors the long scores; and Figure 6C for Program Leadership, which heavily favors the short scores. Thirty such figures were produced for these distributions, and they fell into three groups: a large set where most items clearly favor the long scoring, a small set where some favor the short scoring, and a final set where the two were about equal.

Summary statistics are in Table 2, in which the correlation coefficient reflects the amount of scatter, the slope suggests the average long-short difference, and the paired t-tests confirm the long-short disparities.

In these distributions the ideal result is that the correlation is high (little scatter), the mean difference is zero, and the slope is near unity. Then all points in the scattergram fall near the one-to-one line from zero at the lower left to four at the upper right. When a slope is far from unity the short scores depart systematically from the long scores. For global comparisons a small correlation is acceptable if the slope is near unity and the means are the same, but a small correlation would indicate large disparities across countries.

Fortunately, smaller long-short differences (slope near unity) tend to accompany higher long-short correlations, indicating a consistency across most countries. For example in Figure 4, the top half of items average a 0.79 correlation, while the bottom half average 0.52.

Then which of the 30 items show the best agreement? In Figure 4 the top 15 items contain five of the six Method Availability items, and eight of the thirteen Services items with one more falling just below the top 15. Only one of the three Evaluation items is in the top group. Finally, agreement is least for the Policy items: only one of eight is in the top group.

What explains these patterns? Not regions, since they differ little in their patterns of long-short differences. Not the average level of program effort, since the difference patterns were essentially the same for the high scorers and the low scorers. The most reasonable explanation is that the long-short differences emerge from disparities in the question wording, and in the codes used for the long questionnaire. Where the long and short approaches are closest the scores agree best. An example concerns the close agreement for the six Availability items: the long questionnaire asks for simple percentages and the short questionnaire asks for marks on a simple 1 to 10 scale. Conversely, agreement is less in a few cases where the long questionnaire item contains many cells for the responses (such as training) and checks in numerous cells tend toward high scores, well above the short results. Agreement is fairly close for the budget score because the wordings are simple and similar, but agreement would be even better except that the coding rules for the long item give a zero score for any response below 50% of funds from local sources. (We have kept that rule over the various cycles to retain a consistent time trend in each country.) Thus coding rules for the long form can also play a part. On the other

hand, the short questionnaire has a single format and a single coding plan throughout; however the wording chosen in the brief descriptions of the items can unintentionally invite either lower or higher scores.

Predicting the Long Scores

If the short scores were sufficiently close to the long scores they could be used alone, with savings on cost and time while still protecting the time trend. Table 2 confirms that the long-short correlations are high for most of the 30 scores, in the context of a systematic average difference between the two measures. That suggests that a calibration factor can be obtained, to estimate the long scores from the short ones.

We explored this further by relating the two total scores in a random half of the countries, and applying the equation to the other half. Forty-four of the 88 countries were selected by taking every other one in a list of the countries ordered alphabetically within each region. Among these countries a linear equation between the long and short scores was selected since it gave as good or better fit than several alternatives. When this equation was applied to the other half of countries, to predict the long scores from the short ones, it showed very close agreement. Out of a maximum score of 120 (30 features times 4), most discrepancies were small, around an average difference of only 5.5 points. That concerns the two sets of total scores; the results would be less impressive for the 30 individual scores.

Nevertheless there remains the possibility of using only the short scores in a future cycle of the study. Calibration factors from 1999 could be used to adjust future short scores to what the long ones would have been. That might prove attractive for scores at a certain level of generality, for the total score or for the four components or for other score groupings, but calibration does not appear able to produce a reliable estimate for each of the 30 items in each country individually.

Discussion and Conclusions

The invention of the long questionnaire, and the invention of the brief phrasings in the short questionnaire, both started from a common point – the description of each score that Lapham and Mauldin developed (Appendix B). However, in the long form the respondent's mental focus is upon the particulars of the many questions, not upon the concept of each score as described in the short form. Formats and codes in the long questionnaire matter; and it is probable that a few long scores come from questions in which many sub-

parts can be marked positively, thus yielding high scores, or from codes that tend toward either high or low outcomes, and that in the short form some phrases are worded severely, yielding low scores.

The motivation of the study was largely to see how well the short results would parallel the long results. Simply put, the short scores fall modestly below the long ones, and the seriousness of the differences "depends." The modest disparity between the two outcomes is very consistent in the overall averages for the four components and for the various geographic regions, and it is surprisingly modest even in the country averages for the 30 items. However when individual countries are examined for individual items, there is substantial disagreement between the long and short scores in many cases.

Had the 1999 cycle of the FPE study dropped the long form and used just the short form, that would have shown no gain in the average total score across the developing world from 1994 to 1999, whereas in fact the long form found a continuing improvement. The issue remains open as to what to do in the next cycle, whenever it occurs. Protecting the time trend calls for the long form, but each cycle constitutes a large undertaking. Meanwhile specialized studies of effort have appeared for HIV/AIDS and for maternal and neonatal health, as well as an examination of the policy environment for five reproductive health programs. (All of these rely upon the judgments of expert observers in each country, an approach that is quite distinct from sample surveys.) By the time another cycle of the study is under consideration the various alternatives will have become clearer.

REFERENCES

Lapham, R.J., and W.P. Mauldin. 1972. "National family planning programs: Review and evaluation." *Studies in Family Planning* 3: 3, 29-52.

Lapham, R.J., and W.P. Mauldin. 1984. "Family planning program effort and birthrate decline in developing countries." *International Family Planning Perspectives* 10,4: 109-118.

Mauldin, W.P., J.A. Ross, J. Kekovole, B. e-Khuda, and A. Barkat. 1995. "Direct and judgmental measures of family planning program inputs," *Studies in Family Planning*, 26,5: 287-295.

Mauldin, W.P., and J.A. Ross. 1996. "Family planning programs: Efforts and results, 1972-94. *Studies in Family Planning* 27,3:137-147.

Ross, J.A., and J. Stover. 2000. "Effort Indices for national family planning programs, 1999 cycle." Manuscript; submitted for publication.

Ross, J.A., and W.P. Mauldin. In press. "Measuring the strength of family planning programs." Prepared for the IUSSP/EVALUATION Seminar on Methods for the Evaluation of Family Planning Program Impact. Jaco, Costa Rica, May 14-16, 1997.

Figure 1. Short Total Score vs. Long Total Score, as Percent of Maximum, 88 Countries

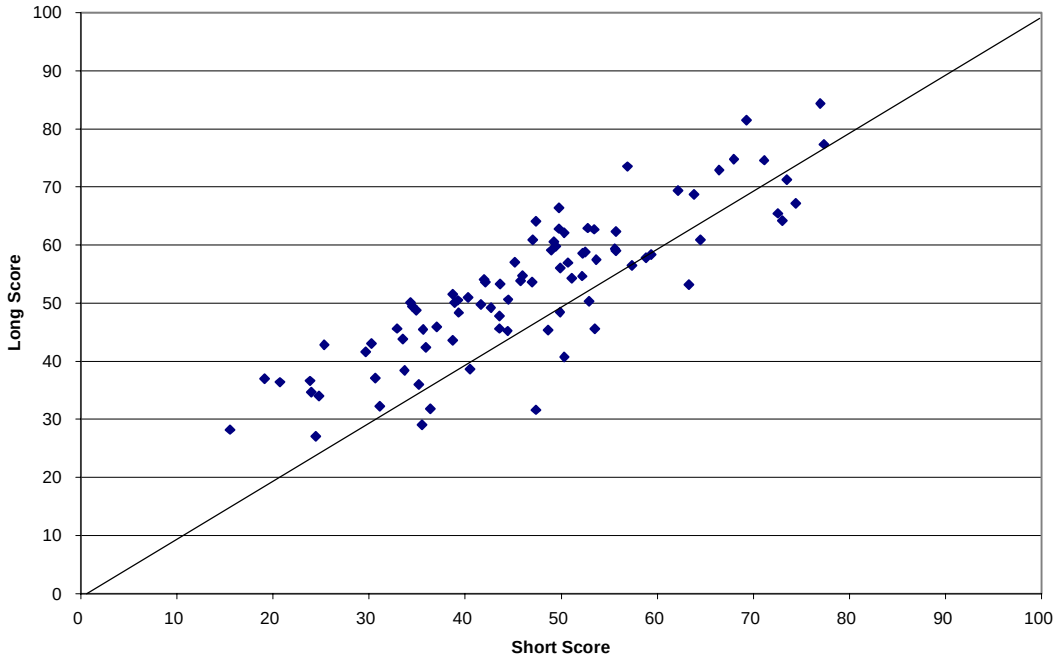

Figure 2. Long vs. Short Program Effort Scores

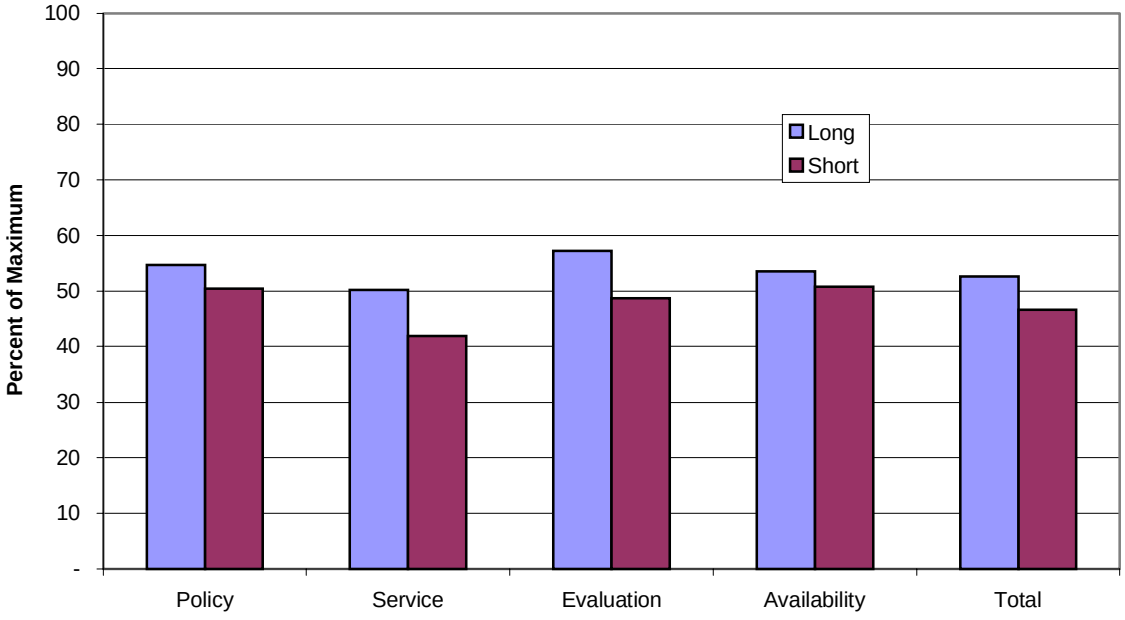

Figure 3. Long versus Short Total Scores, as Percent of Maximum, by Region, for 88 Countries

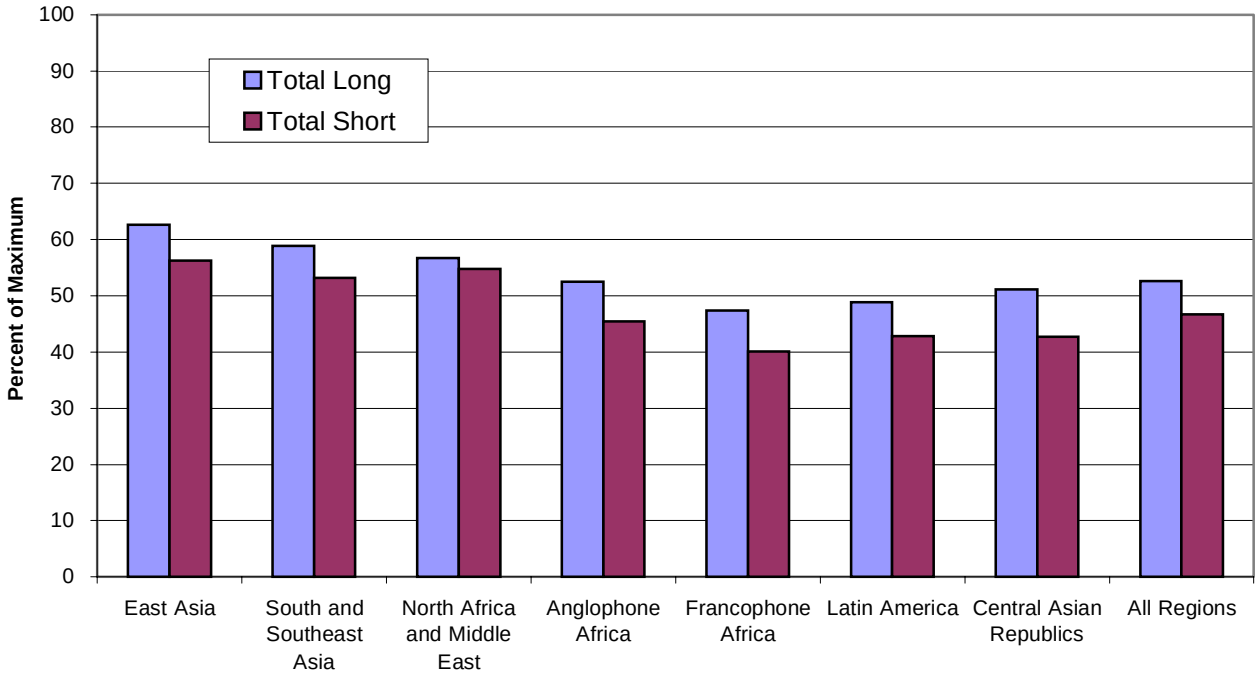

Table 1. FPE Long and Short Scores, Percent of Maximum, by Component and by Region

Region	Policy Long	Policy Short	Service Long	Service Short	Evaluation Long	Evaluation Short	Availability Long	Availability Short	**Total Long**	**Total Short**	**Short as % of Long**	**No. of Countries**
East Asia	57	52	55	50	64	58	85	74	**63**	**56**	90	4
South/South East Asia	60	59	55	47	57	56	65	58	**59**	**53**	90	15
N Africa/ Middle East	60	55	53	51	67	66	56	58	**57**	**55**	97	11
Anglophone Africa	56	51	52	42	59	43	45	46	**53**	**45**	86	15
Francophone Africa	55	45	51	38	57	46	25	36	**47**	**40**	85	16
Latin America	48	47	43.	36	51	41	64	52	**49**	**43**	88	22
Central Asian Republics	50	44	47	38	55	38	60	54	**51**	**43**	84	5
All Regions	55	50	50	42	57	49	54	51	**53**	**47**	89	88

Figure 4. Differences Between Long and Short Scores, 88 Countries

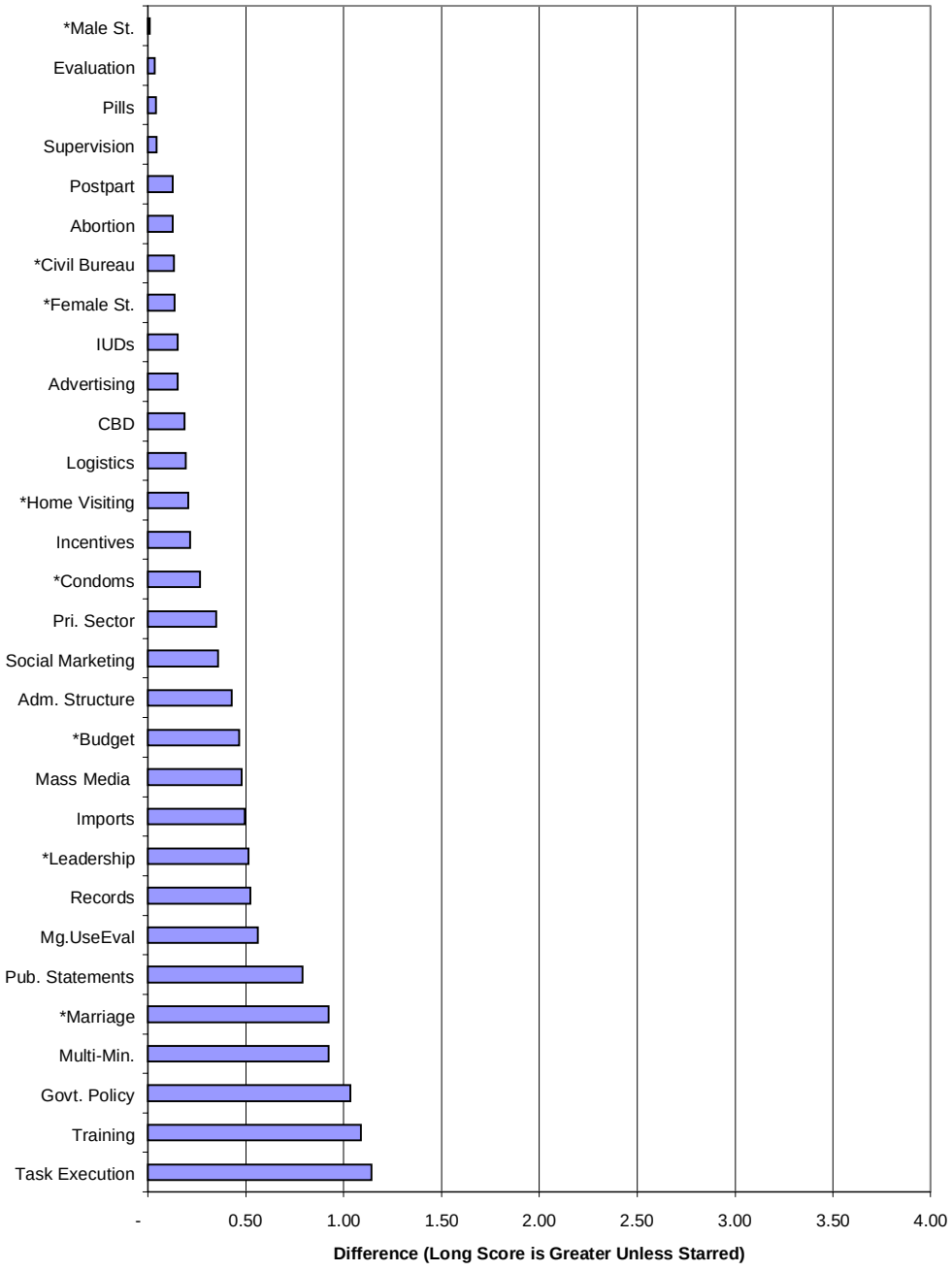

Figure 5. Excess of Long Over Short Mean Scores, 88 Countries

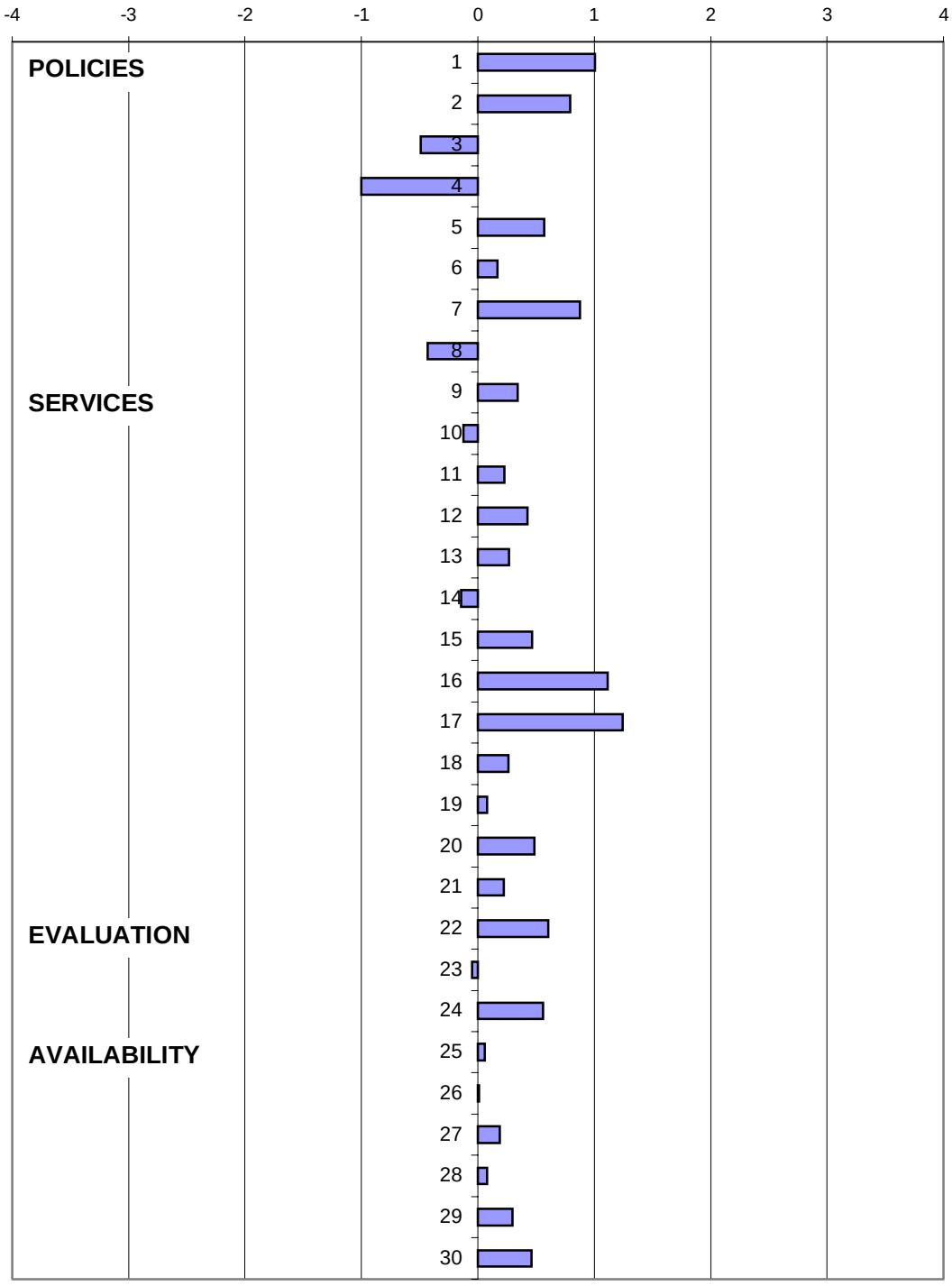

Figure 6A. Female Sterilization: Long versus Short Scores, 88 Countries

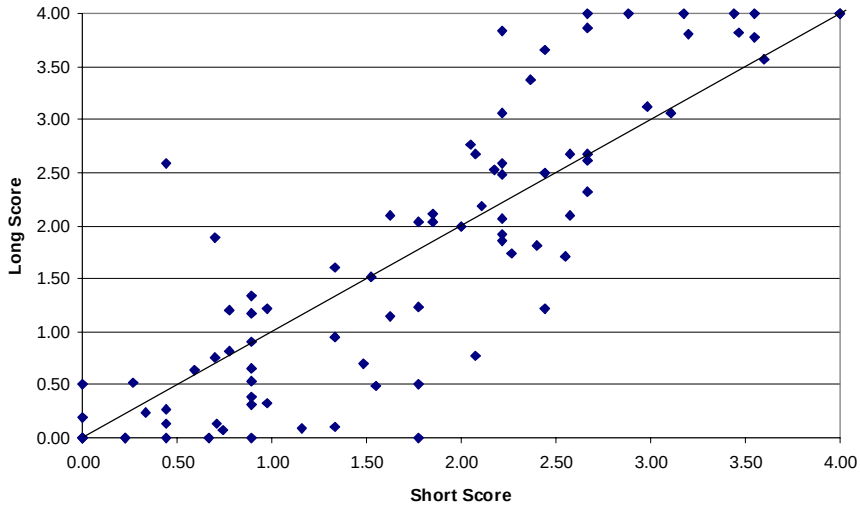

Figure 6B. Training: Long versus Short Scores, 88 Countries

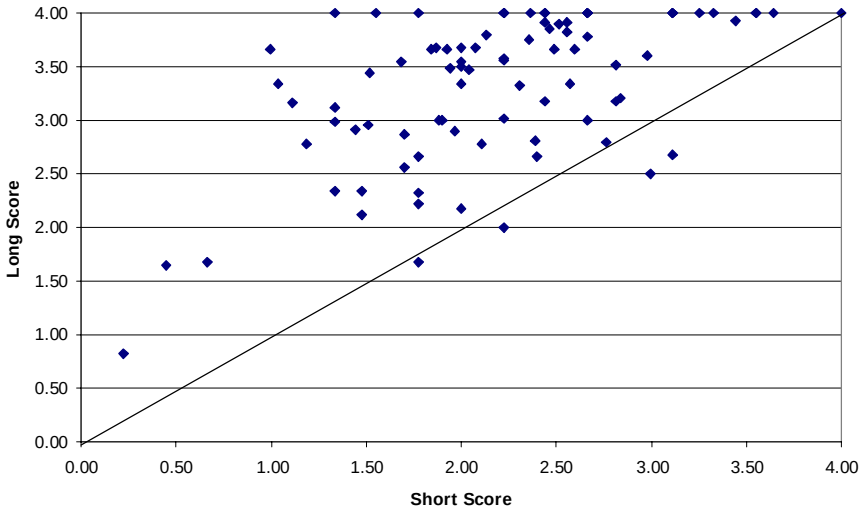

Figure 6C. Program Leadership: Long versus Short Scores, 88 Countries

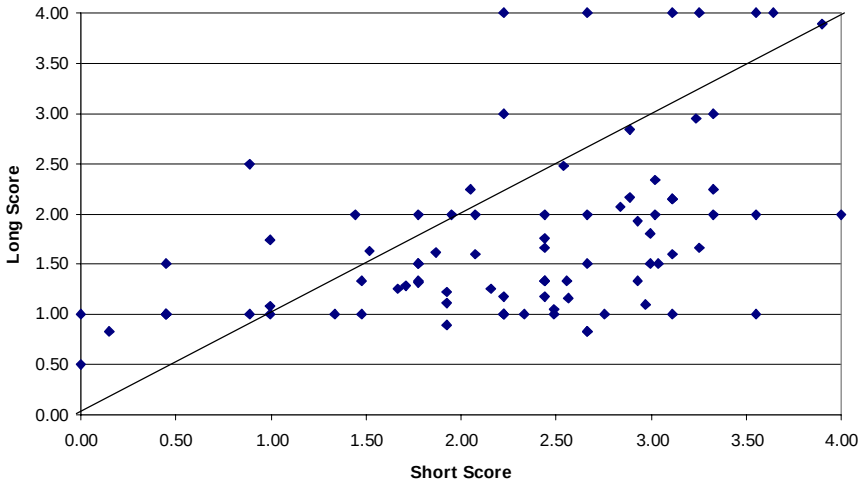

Table 2. Long vs. Short Scores for 88 Countries: Differences, Slopes, Correlations, and Paired T-Tests

Item	Long - Short	Correlation Coefficient*	Slope	Paired T-value	p-value
1	0.99	0.70	0.64	11.51	0.00
2	0.78	0.81	0.94	10.43	0.00
3	(0.52)	0.49	0.45	(5.24)	0.00
4	(0.98)	0.33	0.22	(8.40)	0.00
5	0.57	0.44	0.47	5.59	0.00
6	0.17	0.66	0.87	1.52	0.13
7	0.88	0.57	0.64	9.37	0.00
8	(0.47)	0.91	1.15	(6.23)	0.00
9	0.34	0.34	0.37	3.29	0.00
10	(0.14)	0.68	0.71	(1.74)	0.09
11	0.21	0.87	1.23	3.01	0.00
12	0.40	0.78	1.16	4.00	0.00
13	0.23	0.80	1.09	3.04	0.00
14	(0.15)	0.81	1.13	(1.82)	0.07
15	0.48	0.55	0.70	4.33	0.00
16	1.12	0.56	0.55	15.66	0.00
17	1.23	0.37	0.35	14.61	0.00
18	0.27	0.78	0.81	4.14	0.00
19	0.08	0.68	0.76	1.02	0.31
20	0.48	0.54	0.72	5.25	0.00
21	0.22	0.61	0.71	3.31	0.00
22	0.61	0.80	0.78	10.52	0.00
23	(0.05)	0.61	0.62	(0.60)	0.55
24	0.58	0.58	0.57	7.20	0.00
25	0.07	0.88	1.13	0.95	0.35
26	0.01	0.86	1.10	0.11	0.92
27	0.19	0.56	0.80	1.95	0.05
28	0.08	0.53	0.71	0.94	0.35
29	0.29	0.73	0.96	3.00	0.00
30	0.38	0.73	0.88	3.23	0.00

*All correlation coefficients are statistically significant at .05.
( ) The short score exceeds the long score.
A slope of one means that on average the short and long scores are equal.

Appendix A

The Short Form Format and Illustrative Questions

To give a summary picture of program effort, please rate the following items. Score each item from 1 to 10, where 1 represents non-existent or very weak effort and 10 represents extremely strong effort.

Component	Description	1= Non existent to 10= Extremely strong									
		1	2	3	4	5	6	7	8	9	10
POLICY AND STAGE-SETTING ACTIVITIES											
Policy on fertility reduction and family planning	Extent to which government policy stresses family planning for demographic reasons over health reasons or is simply neutral or opposed.	1	2	3	4	5	6	7	8	9	10
Statement by leaders	Extent to which the head of government, as well as other officials, speak publicly and favorably about family planning at least once or twice a year	1	2	3	4	5	6	7	8	9	10
Level of program leadership	Level of seniority of the director of the national family planning program and whether directly reports to a high level of government	1	2	3	4	5	6	7	8	9	10
Policy on age at marriage	Extent to which legal age at marriage for females is set at 18 years or higher and is enforced	1	2	3	4	5	6	7	8	9	10
Import laws and legal regulations	Extent to which import laws and legal regulations facilitate the importation of contraceptive supplies or extent to which contraceptives are manufactured locally	1	2	3	4	5	6	7	8	9	10
Advertising of contraceptives allowed	Extent of freedom from restrictions on advertising of contraceptives in the mass media	1	2	3	4	5	6	7	8	9	10
Involvement of other ministries and public agencies	Extent to which other ministries and government agencies assist with family planning activities (e.g., delivery of supplies, services, information, education) or other population activities	1	2	3	4	5	6	7	8	9	10

Appendix B

Summary description of the 30 items included in the family planning program effort scale

1. Policy and stage-setting activities

1. Government's official policy or position concerning fertility family planning and rates of population growth
Existence and type of official policy: to reduce the population growth rate, to support family planning activities for other than demographic reasons, to allow private and/or commercial family planning activities in the absence of government-sponsored activity, or to discourage family planning services.

2. Favorable statements by leaders
Whether the head of the government speaks publicly and favorably about family planning at least once or twice a year, and whether other officials also do so.

3. Level of family planning program leadership
Level of the post (person appointed) to direct the national government family planning program, and whether or not the program director reports to the highest level of government.

4. Age-at-marriage policy
Minimum legal age at marriage for females at least 18 years (higher scores for minimum legal ages of 19 and 20+), and the extent of effort to enforce any changes in the law since 1960 regarding legal age at marriage for females. (The score for the latter component is allowed only if the new legal minimum is at least 18.)

5. Import laws and legal regulations regarding contraceptives
Extent to which import laws and legal regulations facilitate the importation of contraceptive supplies that are not manufactured locally, or the extent to which contraceptives are manufactured within the country.

6. Advertising of contraceptives in the mass media allowed
Whether the advertising of contraceptives in the mass media is allowed with no restrictions, whether there are weak restrictions, whether there are social restrictions, or whether there are strong restrictions.

7. Other ministries/government agencies involved
Aside from the ministry or government agency that has primary responsibility for delivering family planning supplies and services, the extent to which other ministries and governmental agencies assist with family planning and/or other population activities. This involvement or assistance may be provided through the public sector or through private-sector family planning programs or population activities, and is classified as follows: assistance with the delivery of family planning supplies and services, assistance in the form of services particular to that ministry, assistance with family planning information and education in specific ways, membership on a council for family planning that meets at least twice annually, moral support and small miscellaneous assistance, no assistance.

8. In-country budget for program
Percentage of the total family planning/population budget available from in-country sources. A top score is given if in-country sources provide 85 percent or more of the budget; no score is given if these sources provide less than 50 percent of the budget.

II. Service and service-related activities

9. Involvement of private-sector agencies and groups
Extent to which private-sector agencies and groups assist with family planning or other population activities. These groups include family planning associations, and so on. The involvement or assistance with family planning and population activities may include the following: delivery of family planning supplies and services, training, family planning information and education, membership in a family planning interagency group that meets at least twice annually, moral support, or other types of assistance.

10. Civil bureaucracy used
Use of the civil bureaucracy of the government to ensure that program directives are carried out, and the extent to which the senior government administrator at the following levels feels responsible for the success of the program: central government level, provincial or state levels, district/governorate/regency/etc. levels, county levels.

11. Community-based distribution (CBD)
Proportion of the country covered by CBD programs for the distribution of contraceptives in areas not easily served by clinics or other service points. Public and/or private CBD systems are included. The essential feature of CBD is that the contraceptive supplies are available upon request within the village, local community, or local residence neighborhood. CBD programs are assumed to be primarily rural; however, a partial extra score is allowed for urban CBD programs.

12. Social marketing
Proportion of the country covered by a social marketing program, that is, subsidized contraceptive sales in the commercial sector. The essential feature of social marketing is that contraceptives are sold at low cost, i.e., a (heavily) subsidized price, through channels easily available to rural or urban residents, such as local shops, pharmacies, or specially created local sales outlets. Some forms of social marketing are called commercial retail sales (CRS) programs. Social marketing programs are assumed to be primarily urban programs; however, an extra score is allowed for rural programs.

13. Postpartum programs
Extent of coverage of new mothers by postpartum programs, which may be hospital or field-based. Most programs are field-based. (1) For hospital-based programs, the score is constructed from the proportion of deliveries in hospitals and maternity centers for which the new mothers are provided a family planning information and education service (by trained female workers), and the proportion of all deliveries in the country that take place in hospitals and maternity centers (often a small proportion); (2) For field-based postpartum programs, the score is constructed from the proportion of women who deliver at home and are offered a family planning information and education service by trained fieldworkers.

14. Home-visiting workers
Proportion of the population covered by a group of workers whose primary task is to visit women in their homes (at least in the rural areas) to talk about family planning and child care. Account is taken of the population that must be covered by each fieldworker; the score for the proportion of the country covered by fieldworkers is deflated if the average population covered by each home visiting worker is more than 15,000.

15. Administrative structure
Whether there is adequate administrative structure and staff at three levels (national, provincial, and county). Adequate means that the administrative structure is sufficient to ensure that plans developed for each level are carried out, that the administrative structure is capable of recognizing and solving problems that cause low performance, and that the administrative structure is able and willing to use existing resources or to call upon

higher administrative levels in obtaining resources needed to carry out plans for the delivery of family planning supplies and services.

16. Training programs
Whether there is an adequate training program for each category of staff in the family planning program: administrative staff, physicians, nurses, paraprofessionals, village-level distributors, fieldworkers/motivators, staff in other ministries and organizations, others. *Adequate* means that the training provides personnel with the knowledge, information, and skills necessary to carry out their jobs effectively, and that facilities exist to carry out the training. The score is determined by the extent to which the training program, for each category of staff, is very good, moderately good, mediocre or poor, or nonexistent.

17. Personnel carry out assigned tasks
Extent to which each category of family planning program staff carries out assigned tasks (task implementation): administrative staff, physicians, nurses, paraprofessionals, village-level distributors, fieldworkers/ motivators, staff in other ministries and organizations, others. The score is determined on the basis of the extent to which each category of staff carries out assigned tasks very well, moderately well, or poorly.

18. Logistics and transport
Extent to which the logistics and transportation systems are sufficient to keep stocks of contraceptive supplies and related equipment available at all service points at all times, at the following levels: central, provincial, county. The score is based on the availability of supplies and equipment: all or almost all of the time, about half to three-quarters of the time, sometimes, or seldom or never.

19. Supervision
Whether there is an adequate system of supervision at all levels. Adequate means that: (a) supervisors exist at all levels of program operations in sufficient numbers to make possible supervisory visits at least once a month at service delivery levels (and quarterly at higher administrative levels); (b) supervisors in fact make such supervisory visits to the work sites of the persons supervised; (c) during these supervisory visits, encouragement, advice, and support are provided to supervised workers, in addition to any necessary checking of operations and records that assist in the evaluation of worker performance; and (d) supervisors carry through on providing/obtaining supplies and services identified as needed during their visits (or at least make serious attempts to obtain these needed supplies and services).

20. Mass media for information, education, and communications (IEC)
Frequency of mass media messages that provide family planning information, including where family planning services are available, and how much of the country is covered by various types of mass media: newspapers, magazines, radio, television, mobile TEC units (films, etc.), billboards and other outdoor media (buses, etc.), traditional types (puppet shows, folk dances, local theater, etc.), other types. The frequency classifications include: at least once a month, sometimes (about once every 3-6 months), infrequently (about once a year or less often), never.

21. Incentive/disincentives
Use of monetary or other incentives for the adoption of family planning. The incentives may be provided to: clients, recruiters, service personnel (including CBD personnel), communities. The disincentives may refer to individuals or to communities, and include regulations or constraints designed to encourage family planning or small family size.

III. Evaluation and record keeping

22. Record keeping
Whether or not there are record-keeping systems for family planning clients at the clinic level, plus a system for the collection and periodic reporting of summary statistics at regional and national levels (that is, numbers of acceptors, supplies distributed, numbers of workers, and so on), and whether or not there is feedback to each reporting unit from regional or national units. The scoring takes into account the existence of good systems as well as their implementation. Feedback refers to reporting back to lower-level units on a regular basis, with progress measured against some standard, such as acceptance or prevalence targets or trends.

23. Evaluation
Whether or not some or all of the following exist (partial score given for each): regular estimation of prevalence levels and trends (annually or quarterly) using program statistics and estimated continuation rates; measurement every two to four years of family planning prevalence levels and trends using data collection methods that are independent from program statistics (such as contraceptive prevalence studies); implementation of operations research studies designed to help program management understand the program, its problems, and potential improvements; professional staff in an evaluation unit who prepare technically correct periodic reports on the program, what it has achieved, etc.; professional staff who interpret and summarize, for program management, national and regional population data collected through censuses, vital registration systems, and surveys (these staff may be directly associated with the program or with other institutions); good coordination, working relationships, and timely sharing of information between the evaluation unit and other units in family planning programs. Partial score is also given for the existence of universities or research institutes in the country that carry out demographic research, family planning research, or population research of other kinds.

24. Management use of evaluation findings
Extent to which the program managers (decision makers) use the research and evaluation findings to improve the program in ways suggested by those findings.

IV. Availability and accessibility of fertility control methods

25. Male sterilization
Whether or not medically adequate voluntary sterilization services for males are legally and openly available, and the percentage of the population that has ready and easy access to such services

26. Female sterilization
Whether or not medically adequate voluntary sterilization services for females are legally and openly available, and the percentage of the population that has ready and easy access to such services.

27. Pills and injectables
Percentage of couples of reproductive age who have ready and easy access to pills through programs other than CBD and social marketing programs. Ready and easy access means that the recipient spends no more than an average of two hours per month to obtain contraceptive supplies and services. Easy access also implies that the cost of contraceptive supplies is not burdensome, i.e., to meet this criterion, a one-month supply of contraceptives should cost less than 1 percent of a month's wages. (If the availability of injectables is higher than that of pills, the data on injectables were used to score this item.)

28. Condoms, spermicides
Percentage of couples of reproductive age who have ready and easy access to condoms, through programs other than CBD and social marketing programs. Ready and easy access is defined as in item 27, above. (If the availability of other conventional contraceptives is greater than that of condoms, the data on those other methods were used to score this item.)

29. IUDs
Percentage of couples of reproductive age who have ready and easy access to IUDs through programs other than CBD and social marketing programs. Ready and easy access is defined as in item 27.

30. Abortion, menstrual regulation
Proportion of the population that has ready and easy access to abortion services, whether or not abortions are legal, but excluded in the scoring is the availability of abortions carried out only under poor conditions.